

Metadaten und Standards

Jensen, Uwe; Katsanidou, Alexia; Zenk-Möltgen, Wolfgang

Veröffentlichungsversion / Published Version

Sammelwerksbeitrag / collection article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:

GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Jensen, U., Katsanidou, A., & Zenk-Möltgen, W. (2011). Metadaten und Standards. In S. Büttner, H.-C. Hobohm, & L. Müller (Hrsg.), *Handbuch Forschungsdatenmanagement* (S. 83-100). Bad Honnef: Bock + Herchen Verlag. <https://nbn-resolving.org/urn:nbn:de:kobv:525-opus-2318>

Nutzungsbedingungen:

Dieser Text wird unter einer CC BY-NC-SA Lizenz (Namensnennung-Nicht-kommerziell-Weitergabe unter gleichen Bedingungen) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier: <https://creativecommons.org/licenses/by-nc-sa/3.0/deed.de>

Terms of use:

This document is made available under a CC BY-NC-SA Licence (Attribution-NonCommercial-ShareAlike). For more information see: <https://creativecommons.org/licenses/by-nc-sa/3.0>

Handbuch Forschungsdatenmanagement

Herausgegeben von
Stephan Büttner, Hans-Christoph Hobohm, Lars Müller

BOCK + HERCHEN Verlag

Bad Honnef

2011

Die Inhalte dieses Buches stehen auch als Online-Version zur Verfügung:
www.forschungsdatenmanagement.de

Die Onlineversion steht unter folgender Creative-Common-Lizenz:

„Attribution-NonCommercial-ShareAlike 3.0 Unported“

<http://creativecommons.org/licenses/by-nc-sa/3.0/>



ISBN 978-3-88347-283-6

BOCK+HERCHEN Verlag, Bad Honnef

Printed in Germany

2.4 Metadaten und Standards

Uwe Jensen, Alexia Katsanidou, Wolfgang Zenk-Möltgen
GESIS – Leibniz-Institut für Sozialwissenschaften

Standardisierte Metadaten sind eine notwendige Voraussetzung für die Dokumentation und dauerhafte Sicherung von Forschungsdaten. Als Werkzeug fördern sie nachhaltig die Erschließung und Nutzung datenbasierter Forschungsergebnisse.

Metadaten sind Daten oder Informationen, die in strukturierter Form analoge oder digitale Forschungsdaten (Objekte) dokumentieren. Sie beschreiben, erklären, verorten oder definieren Objekte, Ressourcen und Informationsquellen für die Wissenschaft. Hierdurch helfen sie, Forschungsdaten zu managen, zu erschließen, zu verstehen und zu benutzen (NISO, 2004).

Im Folgenden werden Forschungsdaten vereinfacht als das Objekt verstanden, das von einer übergeordneten Ebene aus mit Metadaten beschrieben wird. Dabei ist die Heterogenität von Daten und Metadaten zwischen den wissenschaftlichen Disziplinen und auch innerhalb von Fachdisziplinen für die Standardisierung eine nicht zu übersehende Herausforderung. Auch deshalb erscheint es in disziplinübergreifenden Kontexten besonders sinnvoll, den fachspezifischen Kontext der Datenentstehung und die Charakteristika dieser Forschungsdaten möglichst eindeutig herauszustellen. Dabei stellt das Grundmuster, Daten und Metadaten entlang des Forschungsdatenzyklus zu organisieren (UK Data Archive, 2011), eine wichtige verbindende Klammer dar. Dieses Prinzip ist für die systematische Auswahl und Anwendung von Metadaten erforderlich, denn die Anzahl der Metadatenstandards ist überwältigend und ihre Beziehungen untereinander verkomplizieren die Sache weiter (Riley, 2009a).

2.4.1 Studiendesign und Erhebung von Forschungsdaten

Die ganzheitliche Planung einer Studie ist wichtig, um alle relevanten Ebenen, wie z. B. Forschungsfrage, Theorierahmen, Forschungsstrategie, Design der Datenerhebung und Analyse der Forschungsdaten zu berücksichtigen. Dabei spielen Metadaten und Standards eine besondere Rolle, weil sie maßgeblich dazu beitragen, die Qualität der Studie zu sichern. So definiert etwa ISO 20252:2006 einheitliche Qualitätskriterien in der Markt-, Meinungs- und Sozialforschung. Die konkrete Projektplanung ist oftmals in spezifische Qualitätssicherungssysteme eingebunden. So werden Daten der amtlichen Statistik durch Qualitätsstandards abgesichert und durch Qualitätsberichte dokumentiert (Körner & Schmidt, 2006). Gleichzeitig entwickelt die wissenschaftliche Praxis Empfehlungen und Standards zur Dokumentation von Studien und Daten (ICPSR, 2009a). Dazu zählen die Dokumentation der Studie als Ganzes durch

eine detaillierte Studienbeschreibung, die Dokumentation der Verfahren und Instrumente der Datenerhebung sowie der Charakteristika der Samples (Studienebene), die Dokumentation der Struktur und Eigenart der Datensätze (Datensatzebene) sowie die Beschreibung der Variablen einer Datendatei (Variablenebene).

Darüber hinaus werden disziplinspezifische Metadaten zur **Instrumentenentwicklung** bzw. zur Durchführung der Messung in entsprechenden Berichten oder Protokollen festgehalten. So ist die Entwicklung eines sozialwissenschaftlichen Fragebogens bereits ein langer Prozess, während dessen mehrere Versionen erstellt werden, bevor das Messinstrument erprobt und im Feld angewendet werden kann. Die Dokumentation solcher Versionierungsprozesse und der Kommentare der Primärforscher durch entsprechende Tools (Hopt et al., 2010) sind als Metainformationen etwa bei methodischen Fragen zur Datenqualität sehr hilfreich. Gleiches gilt etwa für die Entwicklung und Standardisierung von Antwortskalen, die umfangreiche Tests benötigen, bevor sie endgültig geeicht sind und z. B. in Skalenhandbüchern (ZIS, 2010) oder in umfangreichen technischen Berichten (PISA, 2006) dokumentiert werden.

Die **kulturellen Kontexte** und Fragen der Mehrsprachigkeit stellen Primärforscher in international vergleichenden Einstellungsuntersuchungen vor besondere Herausforderungen. Hier sind die verschiedenen Facetten der Studien- und Datendokumentationen je Sprache und die integrierten Datensätze durch umfangreiche Metadaten qualitativ hochwertig, adäquat und nutzerfreundlich (Jinfang Niu, 2009) zu dokumentieren. Dazu zählen auch ausführliche Beschreibungen z. B. der länderspezifischen Stichprobenziehung, der Übersetzungsprozesse während der Fragebogenentwicklung und entsprechender Pretests (Survey Research Center, 2010).

Metadaten aus der Phase der **Datenerhebung** beschreiben einen z. T. nicht reproduzierbaren Kontext, der sehr bedeutsam für die Auswertung und Interpretation der Daten ist. Hierzu gehören auch Paradata, die den Prozess der Datenerhebung selbst dokumentieren. Dieses können automatisch generierte Daten von Computer-unterstützten Umfragen (z. B. *call record data* oder *keystrokes*) oder auch Informationen sein, die von Interviewern oder anderen technischen Systemen erfasst wurden (z. B. *digital audio recording*) (Couper, 1998). Ein typisches Beispiel für Paradata sind „*call records*“, die Datum und Uhrzeit eines Kontakts und das Ergebnis (z. B. Interview, Ablehnung) erfassen. Paradata erlauben eine unabhängige Evaluation des Nichtstichprobenfehlers in Umfragen. Sie können Nonresponse-Korrekturen verbessern und dabei auch die Wahrscheinlichkeit voraussagen, von einer kontaktierten Person ein Interview zu erhalten (Kreuter et al., 2010).

2.4.2 Auswertung von Forschungsdaten

Die erste Datenauswertung wird üblicherweise von den Primärforschern einer Studie durchgeführt, die dazu ihre umfangreichen informellen Kenntnisse über die erhobenen Daten nutzen können. Sekundäre Nutzer hingegen brauchen formale Metadaten für die Beschreibung der Daten, um die Daten verstehen zu können. Daher sollten auch Datenanalysen durch entsprechende Metadaten nachvollziehbar abgesichert und dokumentiert werden. Die Art und Weise, wie Daten analysiert werden und welche Analysemethoden dabei eingesetzt werden, stellen notwendige Informationen über den Analyseprozess dar. Insbesondere sollte die Kodierung der Felddaten in der statistischen Analyse etwa durch Syntaxfiles als eine Form von Metadaten betrachtet werden, die u. a. auch die Qualität der Analysen dokumentiert und im Sinne guter wissenschaftlicher Praxis entsprechende Replikationen erlauben.

In diesem Zusammenhang besitzt die Dokumentation der Prozesse zur **Harmonisierung** von Daten einen besonderen Stellenwert. Datenharmonisierung bedeutet die Transformation von Daten aus unterschiedlichen Quellen in standardisierte Maße, die eine vergleichende Auswertung erlauben. Harmonisierung bedeutet meistens auch die Herstellung von „*conversion keys*“ (Umrechnungsschlüssel), mit denen Werte eines oder mehrerer Quellvariablen in neue Werte einer standardisierten Zielgröße transformiert werden. Beispiele aus den Sozialwissenschaften sind Bildungs- und Einkommensvariablen, die unterschiedliche Ausprägungen in verschiedenen Ländern haben. Ohne Harmonisierung könnten vergleichende Analysen hierzu nicht durchgeführt werden. Dieser Prozess sollte dokumentiert und die Metadaten sollten verfügbar gemacht werden (Quandt et al., 2009). Metadaten zur Erhebung und Analyse von Forschungsdaten, auch in Form von Computercode, werden in Zukunft an Bedeutung gewinnen, etwa wenn Datensätze als eigenständige Publikation veröffentlicht werden (Hanson et al., 2011). Gleichzeitig stellen sie einen Teil der Informationen bereit, um die Daten nachvollziehbar für (Re)-Analysen zu nutzen. Dabei sind internationale transdisziplinäre Forschungsprojekte mit noch weitergehenden Anforderungen und Problemen konfrontiert, z. B. wenn anhand langer Zeitreihen und heterogener Datenbestände aus unterschiedlichen Fachdisziplinen ein hochkomplexer Forschungsgegenstand untersucht werden soll. So beschreibt etwa der *Study Implementation Plan* des 2006 von ESSP (*Earth System Science Partnership*) und WHO (*World Health Organisation*) initiierten Projektes „*Global Environmental Change and Human Health*“ eine Reihe von kritischen Schwachstellen bei der Nutzung von vorhandenen Datenbeständen (Confalonieri & McMichael, 2007).

Es bleibt festzuhalten, dass Metadaten über das Studiendesign und den Prozess der Datenerstellung und ihre weitere Aufbereitungsgeschichte oft lückenhaft oder unstrukturiert gesammelt oder veröffentlicht werden, wie etwa eine Umfrage von Bose und Frew (2005) zeigt. Auch deshalb ist ein strukturiertes

Datenmanagement für Forschungsdaten generierende Projekte wichtig, weil sie damit die verschiedenen Aspekte eines verantwortungsvollen Umgangs mit Daten und Metadaten planen, überprüfen und dokumentieren können. Vor diesem Hintergrund haben auch Einrichtungen zur Forschungsförderung damit begonnen, Förderzusagen mit der Erstellung von Datenmanagementplänen zu verknüpfen (NSF, 2010; DFG, 2010).¹

2.4.3 Dokumentation von Forschungsdaten

Für die Dokumentation von Forschungsdaten sind – neben der Beschreibung der Genese der Daten durch ein Projekt und dessen Erhebungsinstrumente – die Bedeutung der Daten selbst zentral. Für die Reproduktion einer Datenanalyse oder die Erstellung von Sekundäranalysen muss die Bedeutung von Messreihen und -werten klar definiert werden. Dabei geht es vor allem um die Bedeutung der oft numerischen, aber manchmal auch alphanumerischen Werte eines Datensatzes. Hier sind die disziplinspezifischen Unterschiede sehr groß, da es um die Dokumentation von Messwerten geht, die sachlich unterschiedliche Bezüge und Einheiten haben. So sind in den Naturwissenschaften Messungen etwa der Temperatur in Grad Celsius oder Grad Kelvin, in der Medizin die Messung des Blutzuckergehalts in Millimol pro Liter oder in Milligramm pro Deziliter bekannt, die zu jeweils unterschiedlichen Interpretationen des Datenwerts führen. Für die Volkswirtschaften gibt es z. B. unterschiedliche Indikatoren für die Messung des Wirtschaftswachstums, wie etwa das nominale und reale Bruttoinlandsprodukt als hoch aggregierte Größen, die auf weiteren Werten beruhen. Hier müssen die Definitionen der gewählten Einheiten sehr klar sein oder durch die Metadaten-dokumentation deutlich gemacht werden. In den Sozialwissenschaften ist es bei der Messung mithilfe von Umfragen nötig, die gestellten Fragen ausführlich zu dokumentieren, da hier kleine Abweichungen große Effekte auf die Antworten haben können. Darüber hinaus ist es üblich, für abstraktere Einstellungen, wie etwa die Meinung zur gegenwärtigen Regierung, Skalen zu bilden. Hierbei ist es für die Interpretation der Werte unerlässlich, eine genaue Dokumentation zur Verfügung zu haben.

Die unterschiedlichen fachspezifischen Anforderungen haben zur Herausbildung verschiedener **Metadaten-Frameworks** geführt, die Inhalte von Metadaten auf verschiedenen Ebenen spezifizieren. Ein Metadaten-Framework kann dabei als ein System sicher ergänzender Standards, Normen und kontrollierter Vokabulare beschrieben werden. Zum Teil handelt es sich um rein inhaltliche Festlegungen für die Metadaten, zum Teil werden bis auf die technische Ebene Formate für die Datendokumentation definiert, wie die Metadaten erfasst und

¹. Vgl. Kapitel 2.3 in diesem Band.

ausgetauscht werden können. Damit werden innerhalb der Disziplinen die Ziele der Interoperabilität und Nachvollziehbarkeit erreicht.

Ein Beispiel aus den Sozialwissenschaften für ein konsistentes Metadaten-system zur Dokumentation von Forschungsdaten ist das DDI Format der *Data Documentation Initiative* (DDI). DDI ermöglicht es, alle Metadaten aus den verschiedenen Stadien des Datenlebenszyklus strukturiert zu erfassen (Vardigan et al., 2008). Forschungsdaten werden mit DDI bezüglich ihrer Konzepte, Erhebung, Verarbeitung, Verteilung, Exploration, Analyse, Wiederverwendung und Archivierung beschrieben. Neben Umfragen können auch komplexe Zeitreihen aus nationalen und internationalen Vergleichsstudien, Indikatoren und andere Aggregatdaten sowie geographische Daten beschrieben werden. Die DDI-Spezifikation ist als XML-Schema (Gregory et al., 2009) definiert und ist kompatibel zu anderen Metadatenstandards, etwa mit ISO/IEC 11179 oder Dublin Core. Durch seine stark modular aufgebaute Struktur und die Verwendung von eindeutig identifizierbaren Elementen, versionierbaren Elementen und ganzen Strukturen, die gepflegt werden können („*maintainables*“), unterstützt DDI vor allem die Wiederverwendung und die Vergleichbarkeit von Metadaten.

Ein komplexes Beispiel für ein Metadaten-Framework aus dem Bereich der Klimaforschung ist das Daten- und Metadatenmodell CERA-2 (*Climate and Environmental Retrieval and Archive*). Hier werden Metadaten von Simulationsdaten zur Klimaentwicklung nach den Regeln des WDC-C (*World Data Center for Climate*) festgelegt, um die Daten aus verschiedenen Instituten einheitlich zu beschreiben und miteinander verbinden zu können. Die CERA-2 Metadaten bestehen aus einer Beschreibung des Experiments und einer Beschreibung des Datensatzes. Dabei werden das Projekt und die beteiligten ForscherInnen sowie der Zeitraum, die erhobenen Variablen und ihre Codierung, sowie Struktur und Format der Daten beschrieben. Zur Beschreibung der Variablen werden die CF Konventionen (*Climate and Forecast Metadata Convention*) eingesetzt, so dass über Standardnamen die Vergleichbarkeit der Daten ermöglicht wird. Die Datensatzbeschreibung nimmt auch Bezug auf Skalen oder Codelisten sowie zu Verfahren zur Erzeugung der Daten und ihrer Qualität. Die Daten erhalten *Persistent Identifier*, werden in einem Langzeitarchiv gesichert und im Fall von Veränderungen durch Errata oder neue Datenversionen ergänzt (Tous-saint, 1999). Aktuell beteiligt sich eine Arbeitsgruppe des für das WDC-C zuständige deutsche Klimarechenzentrums an dem EU geförderten Projekt METAFOR (*Common Metadata for Climate Modelling Digital Repositories*), das sich die Entwicklung eines gemeinsamen Informationsmodells für die Klimawissenschaften zum Ziel gesetzt hat.

Weitere Beispiele für Metadatenspezifikationen sind ISO/TS 17369:2005 (*Statistical Data and Metadata Exchange SDMX*), das weitverbreitet von statistischen Ämtern eingesetzt wird, ISO 19115:2003 für Geographische Information und Dienste sowie ISO/IEC 11179:2003 für Metadaten-Registaturen und

Semantik (Gregory et al., 2009). Aus dem Bereich der Umfrageforschung sind auch die Standards von Triple-S für einen technischen Austausch von Metadaten und die ISO 20252:2006 für die Definition einheitlicher Qualitätskriterien in der Markt-, Meinungs- und Sozialforschung bekannt.

In Ergänzung zum Einsatz internationaler Standards zur Datendokumentation wird zusätzlich die Anwendung von nationalen oder internationalen **Klassifikationssystemen** empfohlen, z. B. ISO Normen für geografische Einheiten (ISO 3166) und Sprachen (ISO 639-1:2002) oder Codes für Berufe (ISCO). Für die Katalogisierung von Forschungsdaten werden entweder fachspezifische oder fachübergreifende Klassifikationen angewendet, etwa die Dewey-Dezimalklassifikation (DDC) zur disziplinübergreifenden Erschließung oder fachspezifische eingesetzt (z. B. *Klassifikation Sozialwissenschaften*, GESIS, 2006).

Darüber hinaus stellen **kontrollierte Vokabulare** einen weiteren wichtigen Baustein zur standardisierten Dokumentation von unterschiedlichsten Objekten und deren Eigenarten bereit. Durch die gezielte Erschließung von Forschungsdaten mithilfe von standardisierten Vokabularen, etwa zur Methodik einer Umfrage (Art, Häufigkeit, Stichprobe der Befragung), wird die Vergleichbarkeit von Datensätzen wesentlich verbessert. Eine besondere Bedeutung haben hier mehrsprachige Vokabulare, wie etwa der mehrsprachige ELSST (*European Language Social Science Thesaurus*), der von europäischen Datenarchiven zur Dokumentation und Erschließung sozialwissenschaftlicher Forschungsdaten (*CESSDA Data Catalogue*) eingesetzt wird. Im Zusammenhang mit der Entwicklung von Datendokumentationsstandards wie dem DDI Standard wird auch die entsprechende Entwicklung und Anwendung kontrollierter Vokabulare vorangetrieben (Jääskeläinen et al., 2009). Die in den Codelisten benutzten Vokabulare werden dazu im Standardformat Genericcode erfasst. Genericcode ist eine Spezifikation des OASIS *Code List Representation TC*, mit der ein Standardmodell und die Präsentation in XML für kontrollierte Vokabulare bereitstellt wird. Eine weitere Möglichkeit der Repräsentation kontrollierter Vokabulare aus dem Bereich Semantik Web ist das *Simple Knowledge Organization System* (SKOS), das im August 2009 als neuer Standard durch das W3C veröffentlicht wurde.

Für die Erstellung umfangreicher Metadatendokumentationen ist es häufig unerlässlich, über **Managementsysteme** oder Software zur Pflege der Metadaten zu verfügen. Diese helfen bei der Eingabe, Verwaltung und Versionierung der Metadaten. Solche Anwendungen sind meist stark an die Bedürfnisse des Fachgebiets angepasst und benutzen proprietäre Datenformate. Für die **Interoperabilität** von Daten und Metadaten sollten die Anwendungen jedoch Standardformate verwenden, die nicht proprietär sind und die einfache Wiederverwendung von Forschungsdaten ermöglichen. Solche Standards sollten als offene Standards von der wissenschaftlichen Gemeinschaft getragen werden, so wie etwa der DDI Standard in den Sozialwissenschaften oder die *Network Common Data Form* (NetCDF), die in der Erdsystemwissenschaft angewendet wird.

2.4.4 Publikation, Zitation und Recherche von Forschungsdaten

Für die Publikation von Analyse-Ergebnissen auf der Basis von Forschungsdaten werden in der Regel Zeitschriftenaufsätze oder Monographien erstellt. Diese werden mit den bekannten Metadaten für Literatur dokumentiert, für die schon lange Standards wie DublinCore, MARC (*Machine-Readable Cataloguing*), MODS (*Metadata Object Description Schema*), METS (*Metadata Encoding & Transmission Standard*) oder Z39.50 (*Information Retrieval Protocol*) etabliert sind.

Für die Verknüpfung der Publikationen mit den verwendeten Forschungsdaten haben sich bereits einige Initiativen gebildet, die eindeutige und dauerhafte **Identifikatoren** (*Persistent Identifiers*, PI) für digitale Daten vergeben. Als technische Lösungen haben sich verschiedene Systeme etabliert, etwa das Handle System der CNRI (*Corporation for National Research Initiatives*), oder das darauf aufbauende DOI System der *International DOI Foundation*. Weitere PI Systeme sind das URN System, welches durch die IETF (*Internet Engineering Task Force*) definiert ist und für den Namensraum URN:NBN durch Nationalbibliotheken verwendet wird, sowie PURL (*Persistent Uniform Resource Locators*) getragen vom OCLC (*Online Computer Library Center*) und Zepheira.

Die bereits für Literatur benutzten und weit verbreiteten DOI Namen werden von der Initiative DataCite verwendet, um auch Forschungsdaten eindeutig und dauerhaft identifizierbar und zitierbar zu machen. DataCite hat 19 Mitgliedsinstitutionen aus Europa, Nordamerika, Asien und Australien, die es sich zum Ziel gesetzt haben, den Datenpublikationsagenten in ihren Fachbereichen die Vergabe von DOI Namen für Forschungsdaten zu ermöglichen. Neben der wissenschaftlichen Replikationsmöglichkeit der Analysen steht bei DataCite das Ziel im Vordergrund, der Leistung von Wissenschaftlern bei der Herstellung von Forschungsdatensätzen durch die Zitierfähigkeit höhere Sichtbarkeit zu verleihen. Dafür vergeben die datenhaltenden Einrichtungen DOIs und verpflichten sich zur Pflege der Metadaten. Dabei wird die DOI mit einer URL verbunden, so dass die zitierten Datensätze schnell über das Web aufgefunden werden können. Diese Trennung von *Persistent Identifier* und dem eigentlichen Ort des digitalen Objekts ermöglicht die langfristige Stabilität der Zitation. Gleichwohl sichert es auch die Möglichkeit des Zugriffs auf die Daten, solange der verantwortliche Publikationsagent seiner Pflicht nachkommt die Metadaten aktuell zu halten. Einschränkungen des Zugriffs, deren Gründe auf Seiten des Publikationsagenten vorliegen (z. B. Datenschutz, Beschränkung auf wissenschaftliche Verwendung, Gebühren), können weiterhin vom Publikationsagenten angewendet werden.

In Deutschland sind Mitglieder von DataCite die Technische Informationsbibliothek Hannover (TIB), die Deutsche Zentralbibliothek für Medizin

(ZBMED), die Deutsche Zentralbibliothek für Wirtschaftswissenschaften (ZBW) und GESIS – Leibniz-Institut für Sozialwissenschaften. GESIS betreibt mit dem Projekt *da|ra* (Registrierungsagentur für sozialwissenschaftliche Daten) bereits ein Portal, in dem Datenanbieter DOI Namen für ihre Daten registrieren können und die Metadaten für interessierte Wissenschaftler zur Verfügung stellen. Diese Metadaten enthalten neben den für die Zitation wichtigen Elementen auch inhaltliche und formale Beschreibungen des Forschungsdatensatzes, die Datenerhebung und beteiligte Personen und Institutionen. GESIS hat dazu das von DataCite entwickelte Metadatenschema um spezifische Elemente für die Sozialwissenschaften ergänzt und diese auch mit dem für sozialwissenschaftliche Metadaten etablierten Standard DDI abgeglichen. In diesen Metadaten kann recherchiert werden und die Datenverfügbarkeit (direkter Download, Datenbestellung oder nur Vor-Ort Nutzung) wird angezeigt. Ziel ist es, einen möglichst direkten Zugang zu den Forschungsdaten zu bieten. In Zukunft wird durch eine Kooperation von GESIS mit der ZBW *da|ra* als einheitliches Portal für die Daten der Sozial- und Wirtschaftswissenschaften zugänglich sein.

In der Zukunft werden Metadatenbestände (*Repositories*) oder Verzeichnisse von Metadatenbeständen an Bedeutung gewinnen, da sie technisch die Recherche nach Forschungsdaten erleichtern. Der bereits etablierte CESSDA *Data Catalogue* zeigt, dass standardisierte Metadaten für die Recherche in Forschungsdatenbeständen eine große Erleichterung darstellen. Die im DataCite Verbund registrierten Daten des Projekts PANGAEA (*Publishing Network for Geoscientific & Environmental Data*) zeigen auch die Möglichkeit der Verbindung von wissenschaftlicher Literatur und den verwendeten Forschungsdaten durch die Verwendung von persistenten Identifikatoren. Weitere in DataCite geplante Anwendungen sind die Erstellung von Zitationsstatistiken für Forschungsdaten, die einen Anreiz zur Herstellung von Datensätzen mit hoher Daten- und Metadatenqualität bieten, und die Integration der Datenregistrierung in den Arbeitsablauf bei der Veröffentlichung von wissenschaftlichen Artikeln.

2.4.5 Langzeitarchivierung von Forschungsdaten

Metadaten und Standards zur Archivierung von Forschungsdaten haben in der Diskussion über deren dauerhafte Verfügbarkeit einen besonderen Stellenwert.

Die Vereinheitlichung von technischen und organisatorischen Abläufen ist eines von vielen Zielen zur Entwicklung von Standards zur langfristigen Sicherung von Forschungsdaten. Bei der Entwicklung von Langzeitarchiven besitzt das **Referenzmodell** für ein *Open Archival Information System (OAIS)* (CCSDS, 2002) seit langem eine zentrale Leitfunktion. Als ISO-Standard (ISO 14721:2003) anerkannt, stellt das Modell einen offenen Standard bereit, der auf Grundlage eines umfassendes Begriffs-, Struktur- und Organisationskonzeptes beschreibt, welche Aspekte zu berücksichtigen sind, damit ein Archiv OAIS

konform ist. Die Art und Weise, wie ein Langzeitarchiv organisatorisch oder technisch zu etablieren ist und welche Daten oder Metadaten dabei berücksichtigt werden, ist dann für die jeweiligen Anwendungskontexte speziell zu entwickeln. Offen heißt auch, dass der Standard durch internationale Reviewverfahren veränderbar ist oder Anstoß für weitere Entwicklungen geben kann (CCSDS, 2009; Brübach, 2010).

In Anlehnung oder in expliziter Übereinstimmung mit OAIS wurde auf internationaler Ebene auch damit begonnen, Kernanforderungen an vertrauenswürdige digitale Langzeitarchive zu entwickeln (Ten Principles, 2008). Praxisorientierte Normen, Methoden und Werkzeuge zur Beurteilung der **Vertrauenswürdigkeit** von Langzeitarchiven stellen z. B. TRAC (*Trustworthy Repositories Audit & Certification: Criteria and Checklist*), DRAMBORA (*Digital Repository Audit Method Based on Risk Assessment*) und DAS (*Data Seal of Approval*) bereit. Das deutsche Kompetenznetzwerk zur digitalen Langzeitarchivierung NESTOR entwickelte auf Grundlage nationaler und internationaler Arbeitsergebnisse einen *Kriterienkatalog vertrauenswürdige digitale Langzeitarchive* (NESTOR, 2008).

Gleichzeitig greifen Forschungsdatenarchive diese Kriterien auf, indem sie den Grad der Übereinstimmung mit dem OAIS Rahmenkonzept oder weiteren qualitätsorientierten Kriterienkatalogen ausweisen. So zeigt etwa das Langzeitdatenarchiv ICPSR mit ihrem *Digital Preservation Policy Framework* anhand von sieben Kriterien (Übereinstimmung mit OAIS, Administrative Verantwortung, Funktionsfähigkeit, Finanzen, Angemessenheit von Prozessen und Technologien, Systemsicherheit, Prüfverfahren), wie es die Bedingungen für ein vertrauenswürdigen Datenarchiv erfüllt (ICPSR, 2007). Auf diesem Hintergrund entstand auch ein umfassender Leitfaden zur Datensicherung, der für das IHSN (*International Household Survey Network*) entwickelt wurde und sich insbesondere an die Verantwortlichen in Statistikämtern richtet (ICPSR, 2009b).

Aus den Anforderungen der Langzeitarchivierung stellt sich die Frage nach den notwendigen Metadaten sowie deren **Standardisierung** und **Interoperabilität** neu.

PREMIS (*Preservation Metadata: Implementation Strategies*) ist ein international anerkannter Standard von Metadaten für die Langzeitarchivierung. Die aktuelle Version 2 des Datenlexikons beschreibt in Form semantischer Einheiten Informationen, die über das Archivsystem und dessen Management bekannt sein sollten, um den Prozess der Langzeitarchivierung zu unterstützen. D.h. auch, dass PREMIS mit seinen Kernelementen nur eine Teilmenge aller Metadaten zur Langzeitarchivierung beschreibt. Gleichzeitig wird nicht vorgegeben, wie diese semantischen Informationen in technischen Systemen repräsentiert werden sollen. Sie können aber mit den Metadatenelementen des seit 2008 vorliegenden PREMIS XML Schemas verknüpft werden (Library of Congress, 2009). Die PREMIS Managing Agency betreut die Entwicklung des Standards

und stellt Diskussionslisten, umfangreiche Informationen und Empfehlungen zur Anwendung des Standards bereit. Das Projekt LMER (*Long-term preservation Metadata for Electronic Resources*) der Deutschen Nationalbibliothek stellte 2005 ein Metadatenschema zur Langzeitarchivierung auf Grundlage des Metadatenmodells der Neuseeländischen Nationalbibliothek vor.

METS (*Metadata Encoding and Transmission Standard*) ist ein XML Schema basierter Standard zur Dokumentation von Metadaten digitaler Objektsammlungen dient. Die Objekte werden in 7 Hauptabschnitten dokumentiert, z. B. die innere Struktur des Objektes (structural map), die Gruppierung von zusammengehörigen Dateien (file section) und administrative Metadaten (technische, Metadaten, Rechteinformation sowie Metadaten über das Ausgangsmaterial). Gleichzeitig können andere Metadatentypen, z. B. zur Erschließung oder Langzeitarchivierung (PREMIS, Dublin CORE, MARC, MIX, usw.) mit dem METS Dokument verknüpft oder in METS integriert werden. Im Sinne von OAIS kann ein METS Dokument als Informationspaket für die Einlieferung, Archivierung oder Bereitstellung verwendet werden (Library of Congress, 2005). Die Homepage des METS Standard stellt umfangreiche technische Dokumentationen und Projekt- und Anwenderinformationen bereit.

Die Auswahl von tauglichen technischen Formaten für die nachhaltige Archivierung und Bereitstellung von digitalen Forschungsdaten stellt eine dauerhafte Herausforderung für das Management eines Langzeitarchivs dar. Bei der Auswahl geeigneter Formate werden TIFF, XML oder PDF/A als langfristig nutzbare Dateiformate herausgestellt (Library of Congress, 2007). Gleichzeitig entstehen immer neue Anforderungen bei der Migration, Emulation, Konversion oder Kapselung von medienspezifischen Forschungsdaten, die im Zuge der Software- und Hardwareentwicklung oder in neuen Anwendungszusammenhängen gelöst werden müssen. So erfordern etwa Datenanalysen oder Harmonisierungen von Skalen in virtuellen Arbeitsumgebungen auch Lösungen zur Langfristsicherung von neu produzierten Daten oder Metadaten (Dickmann et al., 2010).

Konkrete Strategien zum Einsatz von Metadaten in der Langzeitarchivierung müssen die Eigenarten der fachspezifischen Forschungsdaten berücksichtigen. So werden weitere Standards des Bibliothekswesens, z. B. LMER, MIX, textMD, im Zusammenhang mit METS und PREMIS angewendet (Dappert & Enders, 2010).

Domainspezifische Anforderungen an Metadaten zur Langzeitarchivierung, die nicht von METS und PREMIS abgedeckt werden, stellen z. B. geographische Forschungsdaten dar. Hier etablieren sich disziplinspezifische Initiativen, um nach adäquaten Lösungen für diese Herausforderungen zu suchen (McGarva u.a., 2009). Die Entwicklung von Dateninfrastrukturen mit domainspezifischen Daten wie z. B. im Projekt GENESI-DR (*Ground European Network for Earth Science Interoperations – Digital Repositories*) zeigt die komplexe Vielfalt von

Standardisierungsanforderungen. GENESI-DR berücksichtigt etwa die Datenerschließung, die Interoperabilität von Metadaten, den gesicherten, lizenzbasierten Zugang oder spezifische Archivierungsformate, z. B. SAFE (*Standard Archive Format for Europe*) für Geodaten der ESA. Speziell auf die Community der Meeresforscher ausgerichtet, stellt das Projekt MMI (*Marine Metadata Interoperability*) einen umfangreichen Katalog mit fachspezifischen Vokabularen und Standards bereit, um die komplexe Welt der Metadaten für diese Disziplin überschaubar zu machen.

Angesichts der Vielzahl von Standards (Riley, 2009b) und ihrer unterschiedlichen Zielsetzungen bedarf es praktikabler Anwendungskonzepte, die Metadaten zum Verstehen, Präsentieren und Erschließen von Daten mit den Metadaten zur langfristigen Sicherung und Verfügbarkeit verknüpfen. Erst damit werden die konzeptuellen Ansprüche, den gesamten Lebenszyklus von Forschungsdaten abzudecken, einlösbar.

Erwähnte Standards und Metadaten Schemata – Modelle, Kriterien, Projekte

CCSDS, (2002). 650.0-B-1: *Reference Model for an Open Archival Information System (OAIS). Blue Book. Issue 1*, January 2002. Der Vorschlag wurde als ISO 14721:2003 anerkannt. Online: <http://public.ccsds.org/publications/archive/650x0b1.PDF> [Zugriff am 27.07.2011] und http://www.iso.org/iso/catalogue_detail.htm?csnumber=24683 [Zugriff am 27.07.2011].

CCSDS, (2009). 650.0-P-1.1 *Reference Model for an Open Archival Information System (OAIS)*, Draft Recommended Standard, Issue 1.1 August 2009. Online: <http://public.ccsds.org/sites/cwe/rids/Lists/CCSDS%206500P11/Attachments/650x0p11.pdf> [Zugriff am 27.07.2011].

CERA-2. *Climate and Environmental Retrieval and Archive*. Online: <http://www.mad.zmaw.de/wdc-for-climate/cera-data-model/> [Zugriff am 27.07.2011].

CESSDA *Data Catalogue*. Council of European Social Science Data Archives. Online: <http://www.cessda.org/accessing/catalogue/> [Zugriff am 27.07.2011].

CF. *Climate and Forecast Metadata Convention*. Online: <http://cf-pcmdi.llnl.gov/> [Zugriff am 27.07.2011].

da|ra. *Registrierungsagentur für sozialwissenschaftliche Daten*. Online: <http://www.gesis.org/dara/> [Zugriff am 27.07.2011].

DAS. *Data Seal of Approval*. Online: <http://www.datasealofapproval.org/> [Zugriff am 27.07.2011].

DataCite. Online: <http://www.datacite.org> [Zugriff am 27.07.2011].

- DDC. *Dewey-Dezimalklassifikation. Deutsche Informationsseite*. Online: <http://www.ddc-deutsch.de/index.htm> [Zugriff am 27.07.2011].
- DDI. *The Data Documentation Initiative*. DDI ALLIANCE., Secretariat ICPSR (Interuniversity Consortium for Political and Social Research), Michigan, Ann Arbor. Online: <http://www.ddialliance.org/> [Zugriff am 27.07.2011].
- DOI System. International DOI Foundation. Online: <http://www.doi.org/> [Zugriff am 27.07.2011].
- DRAMBORA. *Digital Repository Audit Method Based on Risk Assessment*. Online: <http://www.repositoryaudit.eu/> [Zugriff am 27.07.2011].
- Dublin Core. *The Dublin Core® Metadata Initiative*. Online: <http://dublincore.org/> [Zugriff am 27.07.2011].
- ELSST. *European Language Social Science Thesaurus*. Online: <http://elsst.esds.ac.uk/> [Zugriff am 27.07.2011].
- Genericode. *A standard format for defining code lists*. Online: <http://www.genericode.org/> [Zugriff am 27.07.2011].
- GENESI-DR. *Ground European Network for Earth Science Interoperations – Digital Repositories*. Siehe Nachfolgeprojekt GENESI-DEC (Digital Earth Community) – Technical Info and Tutorials. Online: <http://www.genesi-dec.eu/news/> [Zugriff am 27.07.2011].
- GESIS, 2006. *Klassifikation Sozialwissenschaften*. Online: http://www.gesis.org/download.php?url=/fileadmin/upload/dienstleistung/tools_standards/klass.pdf [Zugriff am 27.07.2011]. Library of Congress, 2005. *METS: Überblick und Anleitung* (Übersetzung: A. Menne-Haritz, Juli 2005). Online: http://www.loc.gov/standards/mets/METSOverview.v2_de.html [Zugriff am 27.07.2011].
- Handle System. CNRI Corporation for National Research Initiatives. Online: <http://handle.net/> [Zugriff am 27.07.2011] und Online: <http://www.cnri.reston.va.us/> [Zugriff am 27.07.2011].
- ICPSR, (2007). *Digital Preservation Policy Framework*. Online: <http://www.icpsr.umich.edu/icpsrweb/ICPSR/curation/preservation/policies/dpp-framework.jsp> [Zugriff am 27.07.2011].
- ICPSR, (2009a). Inter-university Consortium for Political and Social Research, 2009. *Guide to Social Science Data Preparation and Archiving: Best Practice Throughout the Data Life Cycle* (4th ed.). Ann Arbor, MI.. Online: <http://www.icpsr.umich.edu/files/ICPSR/access/dataprep.pdf> [Zugriff am 27.07.2011].

ICPSR, (2009b). Inter-university Consortium for Political and Social Research (ICPSR). 2009. *Principles and Good Practice for Preserving Data, International Household Survey Network*, IHSN Working Paper No 003, December 2009. Online: <http://www.ihnsn.org/home/download.php?file=IHSN-WP003.pdf> [Zugriff am 27.07.2011].

ISCO. *International Standard Classification of Occupations*. International Labour Organisation. Online: <http://www.ilo.org/public/english/bureau/stat/isco/index.htm> [Zugriff am 27.07.2011].

ISO 639-1:2002. *Codes for the representation of names of languages*, Part 1: Alpha-2 code. Online: http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=22109 [Zugriff am 27.07.2011].

ISO 3166. *Standard for country codes*. Maintenance agency. Online: http://www.iso.org/iso/country_codes.htm [Zugriff am 27.07.2011].

ISO 14721:2003. *Space data and information transfer systems -- Open archival information system -- Reference model*. Online: http://www.iso.org/iso/catalogue_detail.htm?csnumber=24683 [Zugriff am 27.07.2011].

ISO 19115:2003. *Geographic information – Metadata*. (Online) http://www.iso.org/iso/catalogue_detail.htm?csnumber=26020 [Zugriff am 27.07.2011].

ISO 20252:2006. *Market, opinion and social research -- Vocabulary and service requirements*. Online: http://www.iso.org/iso/catalogue_detail.htm?csnumber=39339 [Zugriff am 27.07.2011].

ISO/IEC 11179:2003. *Information Technology -- Metadata registries*. Online: <http://metadata-stds.org/11179/> [Zugriff am 27.07.2011].

ISO/TS 17369:2005. *Statistical data and metadata exchange (SDMX)*. Online: http://www.iso.org/iso/catalogue_detail.htm?csnumber=40555 [Zugriff am 27.07.2011].

Library of Congress, 2007. *Sustainability of Digital Formats*. Online: <http://www.digitalpreservation.gov/formats/> [Zugriff am 27.07.2011].

Library of Congress, Caplan, P. Network Development & MARC Standards Office, 2008. *PREMIS verstehen*. Übersetzung: Tobias Beinert, Bayerische Staatsbibliothek. Online: http://www.loc.gov/standards/premis/understanding_premis_german.pdf [Zugriff am 27.07.2011].

LMER. *Long-term preservation Metadata for Electronic Resources*. Online: <http://www.d-nb.de/standards/lmer/lmer.htm> [Zugriff am 27.07.2011].

- MARC. *Machine-Readable Cataloguing*. Online: <http://www.loc.gov/marc/> [Zugriff am 27.07.2011].
- METAFOR. *Common Metadata for Climate Modelling Digital Repositories*. Online: <http://metaforclimate.eu/> [Zugriff am 27.07.2011].
- MMI. *Marine Metadata Interoperability*. Online: <http://marinemetadata.org/conventions/> [Zugriff am 27.07.2011].
- METS. *Metadata Encoding & Transmission Standard*. Online: <http://www.loc.gov/standards/mets/> [Zugriff am 27.07.2011].
- MODS. *Metadata Object Description Schema*. Online: <http://www.loc.gov/standards/mods/> [Zugriff am 27.07.2011].
- NESTOR – Kompetenznetzwerk Langzeitarchivierung, 2008. *NESTOR-Kriterien – Kriterienkatalog vertrauenswürdige digitale Langzeitarchive* (Version II). Frankfurt am Main: NESTOR. Online: <http://nbn-resolving.de/urn:nbn:de:0008-2008021802> [Zugriff am 27.07.2011].
- NetCDF. *Network Common Data Form*. Online: <http://www.unidata.ucar.edu/software/netcdf/> [Zugriff am 27.07.2011].
- OASIS Code List Representation TC. Online: http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=codelist [Zugriff am 27.07.2011].
- PANGAEA Publishing Network for Geoscientific & Environmental Data. Online: <http://www.pangaea.de/> [Zugriff am 27.07.2011].
- PREMIS. *Preservation Metadata: Implementation Strategies*. The PREMIS maintenance activity. Homepage <http://www.loc.gov/standards/premis/> [Zugriff am 27.07.2011].
- PURL. *Persistent Uniform Resource Locators*. Träger OCLC (Online Computer Library Center) und Zepheira <http://purl.org/> – <http://www.oclc.org/> – <http://zepheira.com/> [Zugriff am 27.07.2011].
- Riley J. & Becker D., 2009a. *Seeing Standards: A Visualization of the Metadata Universe*. Online: <http://www.dlib.indiana.edu/~jenlrile/metadatamap/> [Zugriff am 27.07.2011].
- Riley J. & Becker D., 2009b. *Glossary of Metadata Standards*. Online: http://www.dlib.indiana.edu/~jenlrile/metadatamap/seeingstandards_glossary_pamphlet.pdf [Zugriff am 27.07.2011].
- SAFE. *Standard Archive Format for Europe within ESA Earth Observation*. Online: <http://earth.esa.int/SAFE/> [Zugriff am 27.07.2011].

SDMX – *Statistical Data and Metadata Exchange*. SDMX Initiative. Online: <http://sdmx.org/> [Zugriff am 27.07.2011].

SKOS. *Simple Knowledge Organization System*. Online: <http://www.w3.org/TR/skos-reference/> [Zugriff am 27.07.2011].

Survey Research Center, 2010. *Guidelines for Best Practice in Cross-Cultural Surveys*. Ann Arbor, MI: Survey Research Center, Institute for Social Research, University of Michigan. Online: <http://ccsg.isr.umich.edu/pdf/00FullGuidelines3.pdf> [Zugriff am 27.07.2011].

Ten Principles, 2007. *Ten basic characteristics of digital preservation repositories*. The Digital Curation Center (U.K), DigitalPreservationEurope, NESTOR (Germany), Center for Research Libraries (North America). Online: <http://www.crl.edu/archiving-preservation/digital-archives/metrics-assessing-and-certifying/core-re> [Zugriff am 27.07.2011].

TRAC. *Trustworthy Repositories Audit & Certification: Criteria and Checklist*. Online: http://www.crl.edu/sites/default/files/attachments/pages/trac_0.pdf [Zugriff am 27.07.2011].

Triple S. *Survey Interchange Standard*. The Triple-S Group. Online: <http://www.triple-s.org/> [Zugriff am 27.07.2011].

URN:NBN Namensraum durch Nationalbibliotheken z. B. für die Deutsche Nationalbibliothek. Online: <http://www.persistent-identifier.de/?link=3352> [Zugriff am 27.07.2011].

URN System. IETF. Internet Engineering Task Force. Online: <http://www.ietf.org/> [Zugriff am 27.07.2011].

WDC-C. *World Data Center for Climate*. Online: <http://www.mad.zmaw.de/wdc-for-climate/> [Zugriff am 27.07.2011].

Z39.50. *Information Retrieval Protocol (Z39.50/ISO 23950 – ANSI/NISO Z39.50)* Online: <http://www.loc.gov/z3950/agency/> [Zugriff am 27.07.2011].

ZIS 2010. *Zusammenstellung sozialwissenschaftlicher Items und Skalen (ZIS)*, Version 14.00, 2010. Online: <http://www.gesis.org/unser-angebot/studienplanen/zis-ehes/> [Zugriff am 27.07.2011].

Literaturhinweise

- Bose, J. & Frew J., 2005. Lineage retrieval for scientific data processing: a survey. *ACM Computing Surveys*, 37(1), March 2005, S. 1–28. Online: <http://dx.doi.org/10.1145/1057977.1057978>. [Zugriff am 27.07.2011]
- Brübach, N., 2010. Die Überarbeitung und Ergänzung des OAIS. In: H. Neuroth et al., Hrsg. 2010. *NESTOR-Handbuch: Eine kleine Enzyklopädie der digitalen Langzeitarchivierung*. (Version 2.3.) Göttingen. Kap.4:13, Kap.4:15- Kap.4:16. Online: <http://nbn-resolving.de/urn/resolver.pl?urn:nbn:de:0008-2010071949> [Zugriff am 27.07.2011].
- Couper, M., 1998. Measuring survey quality in a CASIC environment. In: *Proceedings of the Section on Survey Research Methods of the American Statistical Association*. Online: http://www.amstat.org/sections/srms/proceedings/papers/1998_006.pdf [Zugriff am 27.07.2011].
- Dappert, A. & Enders, M., 2010. Digital Preservation Metadata Standards. *Information Standards Quarterly* (ISQ), 22(2), Special Issue: Digital Preservation. Online: http://www.loc.gov/standards/premis/FE_Dappert_Enders_MetadataStds_isqv22no2.pdf [Zugriff am 27.07.2011].
- DFG (Deutsche Forschungsgemeinschaft), 2010. *Leitfaden für Antragstellung*. (DFG-Vordruck 1.02 – 8/10. S. 32, Abschnitt 3.7 Umgang mit den im Projekt erzielten Forschungsdaten) Online: http://www.dfg.de/download/programme/emmy_noether_programm/antragstellung/1_02/1_02.pdf [Zugriff am 27.07.2011].
- Dickmann, F. Henke, H. & Harms, P., 2010. *Technische Evaluation der Grid-Technologie für das Modellprojekt. Kollaborative Datenauswertung und virtuelle Arbeitsumgebung – VirtAug*. (SOEB Arbeitspapier 2010-1) Online: http://www.wissgrid.de/publikationen/Expertise_VirtAug.pdf [Zugriff am 27.07.2011].
- Confalonieri, U. & McMichael, A. eds., 2007. *Global Environmental Change and Human Health: Science Plan and Implementation Strategy*. (Earth System Science Partnership (DIVERSITAS, IGBP, IHDP, and WCRP) Report No.4; Global Environmental Change and Human Health Report, No. 1), S. 68–72. Online: http://www.gechh.unu.edu/FINAL_GECHH_SP_UPDATED.pdf [Zugriff am 27.07.2011].
- Gregory, A. Heus, P. & Ryssevik, J., 2009. *Metadata*. Working Paper Series of the Council for Social and Economic Data. (RatSWD Working Paper No. 57) (March 2009) Online: http://www.ratswd.de/download/RatSWD_WP_2009/RatSWD_WP_57.pdf [Zugriff am 27.07.2011].

- Hanson, B. Sugden, A. & Alberts, B., 2011. Making Data Maximally Available. *Science*, 331(6018), S. 649. Online: <http://dx.doi.org/10.1126/science.1203354> [Zugriff am 27.07.2011]
- Hopt, O. et al., 2010. *Questionnaire management and DDI: The QDDs case*. DDI Working Paper Series. Online: <http://www.ddialliance.org/sites/default/files/QuestionnaireManagementAndDDI-TheQDDSCase.pdf> [Zugriff am 27.07.2011].
- Jääskeläinen, T. Moschner, M. & Wackerow, A., 2009. Controlled Vocabularies for DDI 3: Enhancing Machine-Actionability. *IASSIST Quarterly*, 33 (1/2), S 34 -39. Online: http://www.iassistdata.org/downloads/iqvol3312wackerow_0.pdf [Zugriff am 27.07.2011].
- Jinfang Niu, 2009. *Perceived Documentation Quality of Social Science Data*. Ph. D The University of Michigan. Online: <http://deepblue.lib.umich.edu/handle/2027.42/63871>. [Zugriff am 27.07.2011].
Siehe auch
Jinfang Niu, 2009. *Overcoming Inadequate Documentation*. Online: <http://deepblue.lib.umich.edu/handle/2027.42/78325> [Zugriff am 27.07.2011].
- Körner, A. & Schmidt, J., 2006. *Qualitätsberichte – ein neues Informationsangebot über Methoden, Definitionen und Datenqualität der Bundesstatistiken*. Statistisches Bundesamt, Wirtschaft und Statistik 2/2006, Wiesbaden. Online: <http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/DE/Content/Publikationen/Querschnittsveroeffentlichungen/WirtschaftStatistik/AllgemeinesMethoden/Qualitaetsberichte,property=file.pdf> [Zugriff am 27.07.2011]
- Kreuter, F. & Casas-Cordero, C., 2010. *Paradata*. Working Paper Series of the Council for Social and Economic Data (RatSWD Working Paper No. 136) (April 2010) Online: http://www.ratswd.de/download/RatSWD_WP_2010/RatSWD_WP_136.pdf [Zugriff am 27.07.2011].
- Kreuter, F., et al., 2010. Using proxy measures and other correlates of survey outcomes to adjust for nonresponse: Examples from multiple surveys. *Journal of the Royal Statistical Society, Series A*. Abstract Online: <http://onlinelibrary.wiley.com/doi/10.1111/j.1467-985X.2009.00621.x/abstract>. [Zugriff am 27.07.2011]
- McGarva, G. Morris, S. & Janée, G., 2009. *Technology Watch Report 09-01: Preserving Geospatial Data*. Digital Preservation Coalition. Online: http://www.dpconline.org/component/docman/doc_download/363-preserving-geospatial-data-by-guy-mcgarva-steve-morris-and-gred-greg-janee [Zugriff am 27.07.2011].

- NSF (National Science Foundation), 2010. *Dissemination and Sharing of Research Results – Data Management Plan Requirements*. Online: <http://www.nsf.gov/bfa/dias/policy/dmp.jsp> [Zugriff am 27.07.2011].
- PISA, 2006. *Technical Report*. Online: www.oecd.org/dataoecd/0/47/42025182.pdf [Zugriff am 27.07.2011].
- Quandt, M. Agache, A. & Friederichs, M., 2009. How to Make the Unpublishable Public. The Approach of the CESSDA Survey Data Harmonisation Platform. *The 5th International Conference on e-Social Science*. Köln, Deutschland 24.-26. June 2009. Online: www.ncess.ac.uk/resources/content/papers/Quandt.pdf [Zugriff am 27.07.2011].
- Toussaint, F., 1999. *Wissenschaftliches Datenmanagement: Das „CERA-2 Daten- und Metadatenmodell“*. Online: http://www.mad.zmaw.de/uploads/media/9911_ptb_01.pdf [Zugriff am 27.07.2011].
- UK Data Archive, 2011. *Research Data Lifecycle*. Online: <http://www.data-archive.ac.uk/create-manage/life-cycle> [Zugriff am 27.07.2011].
- Vardigan, M. Heus, P. & Thomas, W., 2008. Data Documentation Initiative: Toward a Standard for the Social Sciences. *The International Journal of Digital Curation*, 3(1). Online: <http://www.ijdc.net/index.php/ijdc/article/view/66/45> [Zugriff am 27.07.2011].